Application of two-stage fuzzy set theory to river quality evaluation in Taiwan

Shiow-Mey Liou*, Shang-Lien Lo, Ching-Yao Hu

Graduate Institute of Environmental Engineering, National Taiwan University, 71, Choshan Road, Taipei 106, Taiwan

Abstract

An indicator model for evaluating trends in river quality using a two-stage fuzzy set theory to condense efficiently monitoring data is proposed. This candidate data reduction method uses fuzzy set theory in two analysis stages and constructs two different kinds of membership degree functions to produce an aggregate indicator of water quality. First, membership functions of the standard River pollution index (RPI) indicators, DO, BOD_5, SS, and NH_3-N are constructed as piecewise linear distributions on the interval [0,1], with the critical variables normalized in four degrees of membership (0, 0.33, 0.67 and 1). The extension of the convergence of the fuzzy c-means (FCM) methodology is then used to construct a second membership set from the same normalized variables as used in the RPI estimations. Weighted sums of the similarity degrees derived from the extensions of FCM are used to construct an alternate overall index, the River quality index (RQI). The RQI provides for more logical analysis of disparate surveillance data than the RPI, resulting in a more systematic, less ambiguous approach to data integration and interpretation. In addition, this proposed alternative provides a more sensitive indication of changes in quality than the RPI. Finally, a case study of the Keeling River is presented to illustrate the application and advantages of the RQI.
© 2003 Elsevier Science Ltd. All rights reserved.

Keywords: River pollutant index; River quality index; Fuzzy theory; Fuzzy c-means; Similarity degree; Sensitive analysis

1. Introduction

Creating and maintaining environmental indicators that are scientifically sound but easy for the lay public to grasp is essential when complex environmental quality trends need to be effectively used in developing and communicating environmental public policy [1,2]. The River pollution index (RPI), for example, is used by EPA of Taiwan to explore monitor trends for both planning and day-to-day management of surface water quality for the public. The RPI involves four parameters: dissolved oxygen (DO), biochemical oxygen demand (BOD_5), suspended solids (SS), and ammonia nitrogen (NH_3-N), each of which is ultimately converted to a four-state quality sub-index (1, 3, 6, and 10). The overall index is then divided into four pollution levels (non-polluted, lightly polluted, moderately polluted, and grossly polluted) by averaging the four sub-indices (see Table 1). Aggregated classification indices were developed in the 1970s [3–5], and the adequacy and sensitivity of this classification to subtle changes in water quality has been the subject of considerable recent debate [6,7]. In the latest two decades, there has been much research into indices [8–13]. One possible improved approach is suggested by developments in the area of fuzzy theory [14], which was first developed by Zadeh [15]. In the previous two decades, the theory of fuzzy sets had advanced in a variety disciplines [16,17], including the calculation of quality indices from environmental monitoring data [18–20]. Fuzzy methods are appealing because they are suited to modelling the continuum characteristic of the underlying complex environmental

*Corresponding author. Tel.: +886-936052767; fax: +886-23928821.

E-mail address: d741004@ms.cc.ntu.edu.tw (S.-M. Liou).

Table 1
The classification ranks defined by the river pollution index (RPI)

Items/ranks	Good	Slightly polluted	Moderate polluted	Gross polluted
DO (mg/L)	Above 6.5	4.6–6.5	2.0–4.5	Under 2.0
BOD_5 (mg/L)	Under 3.0	3.0–4.9	5.0–15	Above 15
SS (mg/L)	Under 20	20–49	50–100	Above 100
NH_3-N (mg/L)	Under 0.5	0.5–0.99	1.0–3.0	Above 3.0
Index scores (S_i)	1	3	6	10
Sub-index	Under 2	2.0–3.0	3.1–6.0	Above 6.0

Sub-index $= \frac{1}{4}\sum_{i=1}^{4} S_i$.

interactions that the quality data seek to measure. The fuzzy c-means (FCM) algorithm has proved to be effective in exploratory detection of data structures, and has successfully been applied to a variety of clustering problems [21]. The convergence analysis of the algorithm has attracted a considerable amount of attention since the publication of Bezdek's convergence theory in 1978 [22].

A linkage of the convergence of FCM is used potentiality. It is widely accepted that, according to the conventional FCM algorithm, similarity measurement could provide the answer to represent the degree to which a data set belongs to a particular group. However, in this study the similarity degrees of an object do not only directly infer the objects to a specific quality level, but also provide an intermediate measurement to convert degrees into an overall quality index. The main goal of the present work is that the constituents be transformed and aggregated and an overall index calculated with the area of fuzzy theory. It would be a powerful tool to cope with certain complicated situations. The River quality index (RQI) is referred to in this paper to distinguish it from other overall indices.

2. Materials and methods

2.1. Membership function

A classical (crisp) set is normally defined as a collection of elements that can be a finite and countable space of an object. Each single element can either belong to a set or not. Alternatively, for a fuzzy set one can define the number of elements by using the characteristic function, in which 1 indicates membership and 0 non-membership. [15]. The theory of fuzzy membership function is a theory of graded concepts [23] in which membership in a set is represented as a continuous value rather than the familiar binomial zero or one. Measurements of environmental monitoring parameters encompass natural scales that may range from a few micrograms per litre to hundreds of grams per litre (when the units are a concentration), or may require

comparison among measurements with different natural units. These differences in natural scale between the various quality indicators makes data pre-processing very important because this is a necessary step in allowing an adequate comparison between the different indicators. Poor pre-processing methods can produce biases that may overemphasize the importance of one factor in the overall quality measurement. Subindex was widely used in water quality measurement in earlier research because it recognizes that water quality does not go instantaneously from "good" to "bad" as water quality changes beyond the guideline value [3,8,10,11,24]. In other words, the crucial characteristic (membership function) for fuzzy set theory is the same as the virtue of subindex for quality assessment. The transformation of a parameter estimate into environmental quality is performed through the use of a value function relating the various levels of the parameter estimation to the appropriate levels of environmental quality [10,24].

2.2. Fuzzy c-mean algorithm

The general FC algorithm partitions a data set of n objects or pattern vectors into c clusters or groups ($c \leqslant n$). This partitioning is achieved by minimising an objective function, J_m, using an iterative procedure. The criterion function is as follows.

$$J_m(U, V, X) = \sum_{i=1}^{c} \sum_{k=1}^{n} \mu_{ik}^m \|\vec{x}_k - \vec{v}_i\|_A^2, \ 1 \leqslant m \leqslant \infty, \qquad (1)$$

where $\vec{x}_k \in R^p$ represents an object data k with p-dimension; $X = (\vec{x}_1, \vec{x}_2, ...\vec{x}_n) \in \Re^{p \times n}$ denotes a matrix of object data; the degree of membership, $\mu_{ik} \in [0, 1]$, measures the likelihood of observation $\vec{x}_k$ belonging to cluster i; $U \in \Re^{c \times n}$ is a matrix of similarity degrees; $\vec{v}_i \in R^p$ is the prototype of the ith cluster ($i = 1...c$); $V \in \Re^{p \times c}$ is a matrix of cluster centroids; $\|\vec{x}_k - \vec{v}\|_A^2$ represents distance functions, and when the covariance matrix of all the observations in the data set is equivalent to the identity matrix I, then the distance metric becomes equivalent to the Euclidean distance

norm, and $m \in [1, \infty]$, which is called the fuzziness index, regulates the degree of partition fuzziness. Minimization of J_m under the following constrains yields an iterative minimization pseudo-algorithm well known as the FCM algorithm [25–28].

$$\text{(i)} \quad 0 \leqslant \mu_{ik} \leqslant 1 \ \forall i, k, \tag{2a}$$

$$\text{(ii)} \quad \sum_{i=1}^{c} \mu_{ik} = 1 \ \forall k, \tag{2b}$$

$$\text{(iii)} \quad 0 \leqslant \sum_{k=1}^{n} \mu_{ik} < n. \tag{2c}$$

The components $\vec{v}_i$ and the membership degrees μ_{ik} are updated according to the following expressions.

$$\vec{v}_i = \frac{\sum_{k=1}^{n} (\mu_{ik})^m \vec{x}_k}{\sum_{k=1}^{n} (\mu_{ik})^m}, \ 1 \leqslant i \leqslant c, \tag{3}$$

$$\mu_{ik} = \frac{1/[\|\vec{x}_k - \vec{v}_i\|_A^{21/m-1}]}{\sum_{j=1}^{c} \left[\frac{1}{\|\vec{x}_k - \vec{v}_j\|_A^{21/m-1}} \right]}, \ 1 \leqslant i \leqslant c \ \text{if} \ \|x_k - v_i\|_A^2 > 0, \tag{4a}$$

$$\mu_{ik} = 1, \ 1 \leqslant i \leqslant c, \ \text{if} \ \|x_k - v_i\|_A^2 = 0, \tag{4b}$$

$$\sum_{i=1}^{c} \mu_{ik} = 1. \tag{4c}$$

The FCM algorithm uses Picard iteration through the loop defined by Eqs. (3) and (4) to obtain the prototypes, which produce the minimal J_m for a fixed group number c. An observation, $\vec{x}_k$, is assigned to the cluster i ($i = 1 \ldots c$), when its degree of membership of that particular cluster μ_{ik}, is greater than its membership values of all other clusters. Bezdek provides an excellent treatise on the family of fuzzy k-means methods [29].

2.3. The river quality index model

Based on the convergence theory of the FCM, the similarity measure of two fuzzy sets is revealed in Eq. (4), which explores the relationships between two observations (both vectors are of p-dimensions) related to the values of distance measures. Similarity measures are large when the two objects being compared share a considerable amount of commonality, and small when they differ significantly from each other. Distance measures vary inversely to similarity measures in magnitude [29]. The extension of the above properties is employed in quality assessment. The parameter m regulates the association fuzziness degree of similarity

measures and distance measures. Instead of seeking the optimal prototypes for a fixed group number c, cluster nucleus $\vec{v}_i$ represents quality measurements, which reflects or even is defined as a specific quality level orderly arranged and assigned in advance. The prototypes are defined as the specific quality. Four nucleuses were defined, nil, moderate, severe, and extreme impacts for the fish farm explored in Silvert [30]. Hence, Eq. (4) is redefined as follows in quality evaluating application (see Eq.(5).

$$\mu_{ik} = \frac{1/[\|\vec{f}_k - \vec{e}_i\|_A^{2/(m-1)}]}{\sum_{j=1}^{c} \left[\frac{1}{\|\vec{f}_k - \vec{e}_j\|_A^{2/(m-1)}} \right]}, \tag{5}$$

where $\vec{f}_k$ represents the data point transformed from concentration to the memberships of quality for a given use. The prototype, $\vec{e}_i$, is defined as a specific quality-ordered level and is assigned in advance, and $\mu_{ik} \in [0, 1]$ is the similarity degree of data point $\vec{f}_k$ to ith specific quality level.

The RQI is used to address the monitoring data of the river environment based on this extended FCM methodology. Since the similarity measures measure the commonality of the observation and assigned specific quality-ordered levels, an overall quality index of an observation $\vec{x}_k$, can be obtained from the accumulated summing up of its similarity degrees to all of the specific quality-ordered levels μ_{ik} (for $i = 1$ to c). The greater the commonality of an observation similar to the good quality level, the higher is its overall score gained. It is obvious that only if the increased points are assigned does it not matter how large are the values of the weighting points. The weighting points of quality levels $q_i \in [0, 1]$ (for $i = 1$ to c) are registered into equal parts according to the number of specific quality-ordered levels for having a general formula, which could be applied in any number of quality levels. The points are divided into $q_i = 0$, 0.5 and 1, $c = 3$ for example, which represents the similarity degree weights of the object to three specific quality-ordered levels, respectively. By accumulating one set of weighted similarity degrees, the RQI of observation $\vec{x}_k$ is derived (as Eq. (6)).

$$\text{RQI}_k = (\sum_{i=1}^{c} \mu_{ik} \times q_i) \times 100. \tag{6}$$

Ultimately, for the public's recognition, the factor of 100 is employed. Hence, the value of the RQI is ranged from 0 to 100.

3. Results and discussion

Based on the two-stage fuzzy set, the RQI is established. To undertake this research, a program was
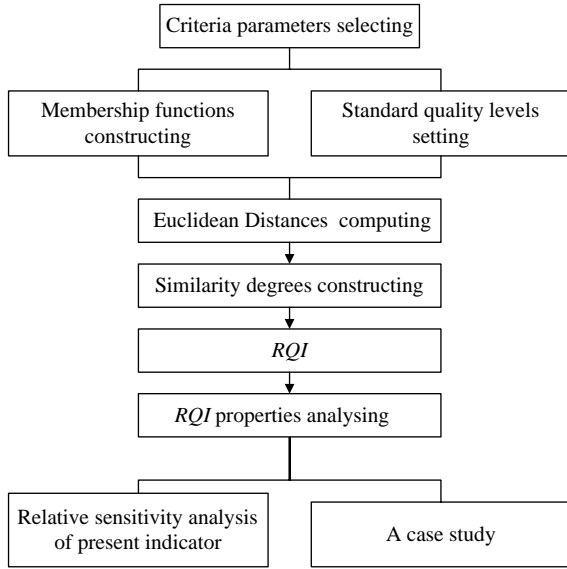
Fig. 1. Schematic procedure of the proposed model for river quality evaluation.

Table 2
The key-points defined in the membership functions

Membership degree (f_x)	0	1/3	2/3	1
DO (mg/L)	Above 6.5	5.5	3.25	Under 2.0
BOD5 (mg/L)	Under 3.0	3.95	10	Above 15
SS (mg/L)	Under 20	34.5	75	Above 100
NH3-N (mg/L)	Under 0.5	0.7	2	Above 3.0

developed to implement the RQI. The flow chart of the program is drawn in Fig. 1.

3.1. Membership functions for quality estimation

The four parameters: DO, BOD$_5$, SS, and NH$_3$-N are employed in the study since these are the key parameters suggested by earlier research [3,7,10,13,27]. Let $X = \{x_d, x_b, x_s, x_n\}$ is a four-dimension sampling space of river water, where d denotes DO (mg/L), b denotes BOD$_5$ (mg/L), s represents SS (mg/L), and n indicates ammonia nitrogen (mg/L). Segmented linear membership functions of the critical variables are constructed. Four crucial breakpoints: 0, 0.33, 0.67 and 1 are registered based on the four specific standard RPI levels. The average concentrations of each parameter in each pollution level of RPI index are determined in accordance with the degree of water quality (see Table 2). The piecewise linear membership functions of the variables

are established as follows.

$$f_d(x_d) = \begin{cases} 1 & \text{for} & x_d > 6.5, \\ 2/3 + (1/3) \times \dfrac{x_d - 5.5}{6.5 - 5.5} & \text{for} & 5.5 \leqslant x_d \leqslant 6.5, \\ 1/3 + (1/3) \times \dfrac{x_d - 3.25}{5.5 - 3.25} & \text{for} & 3.25 \leqslant x_d \leqslant 5.5, \\ (1/3) \times \dfrac{x_d - 2}{3.25 - 2} & \text{for} & 2.0 \leqslant x_d \leqslant 3.25, \\ 0 & \text{for} & x_d < 2, \end{cases} \tag{7}$$

$$f_b(x_b) = \begin{cases} 1 & \text{for} & x_b < 3.0, \\ 2/3 + (1/3) \times \dfrac{3.95 - x_b}{3.95 - 3.0} & \text{for} & 3.0 \leqslant x_b \leqslant 3.95, \\ 1/3 + (1/3) \times \dfrac{10 - x_b}{10 - 3.95} & \text{for} & 3.95 \leqslant x_b \leqslant 10.0, \\ (1/3) \times \dfrac{15 - x_b}{15 - 10} & \text{for} & 10.0 \leqslant x_b \leqslant 15.0, \\ 0 & \text{for} & x_b > 15.0, \end{cases} \tag{8}$$

$$f_s(x_s) = \begin{cases} 1 & \text{for} & x_s < 20.0, \\ 2/3 + (1/3) \times \dfrac{34.5 - x_s}{34.5 - 20.0} & \text{for} & 20.0 \leqslant x_s \leqslant 34.5, \\ 1/3 + (1/3) \times \dfrac{75 - x_s}{75 - 34.5} & \text{for} & 34.5 \leqslant x_s \leqslant 75, \\ (1/3) \times \dfrac{100 - x_s}{100 - 75} & \text{for} & 75 \leqslant x_s \leqslant 100, \\ 0 & \text{for} & x_s > 100, \end{cases} \tag{9}$$

$$f_n(x_n) = \begin{cases} 1 & \text{for} & x_n < 0.5, \\ 2/3 + (1/3) \times \dfrac{0.7 - x_n}{0.7 - 0.5} & \text{for} & 0.5 \leqslant x_n \leqslant 0.7, \\ 1/3 + (1/3) \times \dfrac{2 - x_n}{2 - 0.7} & \text{for} & 0.7 \leqslant x_n \leqslant 2, \\ (1/3) \times \dfrac{3 - x_n}{3 - 2} & \text{for} & 2 \leqslant x_n \leqslant 3, \\ 0 & \text{for} & x_n > 3, \end{cases} \tag{10}$$

where $f_d(x_d), f_b(x_b), f_s(x_s)$, and $f_n(x_n)$ represent the membership functions of parameter DO, BOD$_5$, SS, and NH$_3$-N, respectively. The rating curves of the four criteria variables resulting from membership functions as described above are shown in Fig. 2. Nives [10] states that selection of the rated value of quality is an arbitrary one but the base of the survey and the application of this method to the evaluation of surface water quality can be modified according to their territorial circumstances elsewhere.
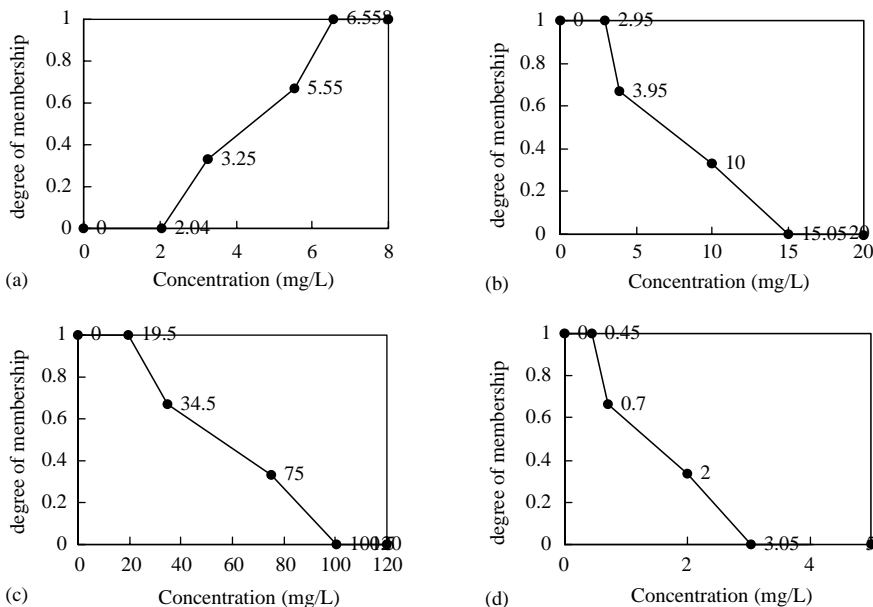
Fig. 2. The rating curves of the employed parameters: (a) DO (mg/L), (b) BOD$_5$ (mg/L), (c) SS (mg/L), (d) NH$_3$-N (mg/L).

Table 3
The notations used in the synthetic data study

Notation	Definition
$\vec{e}_{0/n-1},\ \vec{e}_{1/n-1},\ldots,\vec{e}_{n-2/n-1}$ and $\vec{e}_{n-1/n-1}$	n specific quality-ordered levels
$\mu_{0/n-1},\ \mu_{1/n-1},\ldots,\ \mu_{n-2/n-1}$ and $\mu_{n-1/n-1}$	The similarity among the synthetic data and their identified quality levels
$q_{0/n-1},\ q_{1/n-1},\ldots,\ q_{n-2/n-1}\ q_{n-1/n-1}$	The weighting points coincident with quality-ordered levels

3.2. Optimal fuzziness index, m, and suitable number of specific quality level, c

The properties of the similarity degrees based on the convergence theory of the FCM are studied. Groups of synthetic data sets are used in which each of the observations $\vec{f}_k$ (for $k=1$ to n) consists of four homogeneous values (four-dimensional measurement space). In total, 101 subsets of observations ranging from the terminal values $\vec{f}_0 = (0,0,0,0)$ to $\vec{f}_{101} = (1,1,1,1)$ are used to compute the similarity degrees and the RQI. The usable parameters in Eq. (5) are set as follows: The Euclidean norm is chosen for distance function being well known and commonly used. The fuzziness index, m, is set with 7/5, 5/3, 2, and 3. By varying specific quality-ordered levels, c, from 2 to 5, four sets of similarity degree functions can be derived for each fuzziness index. In total, 16 index curves (4 values of m for $m \times 4$ possibilities for the quality levels) are generated. Table 3 explicates the notations used in the synthetic data study. The subsets of $\vec{e}_0 = (0,0,0,0)$ and $\vec{e}_1 = (1,1,1,1)$, for example, represent the specific

quality-ordered level of perfectly "bad" and perfectly "good" quality. In other words, these represent the absolute poorest measurements and absolute best measurements that are possible. Using the definition of μ_{jk} in Eq. (5), it is possible to calculate the similarity degrees, μ_{0k} and μ_{1k}, between object $\vec{x}_k$ to the two levels from interval [0, 1]. This can be represented as follows (see Eq. (11)).

$$\mu_{0k} = \frac{1/\|\vec{f}_k - \vec{e}_0\|^{2/(m-1)}}{(1/\|\vec{f}_k - \vec{e}_0\|^{2/(m-1)}) + (1/\|\vec{f}_k - \vec{e}_1\|^{2/(m-1)})},$$
$$\mu_{1k} = \frac{1/\|\vec{f}_k - \vec{e}_1\|^{2/(m-1)}}{(1/\|\vec{f}_k - \vec{e}_0\|^{2/(m-1)}) + (1/\|\vec{f}_k - \vec{e}_1\|^{2/(m-1)})}. \quad (11)$$

An overall index of object $\vec{x}_k$ is computed by $\text{RQI}_k = [(\mu_{0k}) \times 0 + (\mu_{1k}) \times 1] \times 100$ referring to Eq. (6). Hence, the subsets of $\vec{e}_0 = (0, 0, 0, 0)$, $\vec{e}_{1/4} = (0.25, 0.25, 0.25, 0.25)$, $\vec{e}_{2/4} = (0.5, 0.5, 0.5, 0.5)$, $\vec{e}_{3/4} = (0.75, 0.75, 0.75, 0.75)$, and $\vec{e}_1 = (1, 1, 1, 1)$, for example, sort the RQI based on five specific quality levels.

The curves of RQI for the 101 synthetic data sets based on two specific quality levels with five varying fuzzy index m are shown in Fig. 3. Essentially, the RQI rises with the increasing value of the defaulted data set, $\vec{f}_k$ so the quality of the observation could be explored. In the case of $c = 2$, $m = 3$, the summarized RQI yields the average of the distance of a data set to a cluster nucleus so that the linear relationship between the subset of synthetic data $\vec{f}_k$ and RQI_k perfectly exists. The slight "S" curves display when either $m = 2$ or $m = 5$. Occasionally, such curves are preferred in quality assessment since they are bonded to the membership functions not being perfect linearity in use. The RQI tend to be crisp when $m = 11/10$ of which the values are either 100 or 0. Furthermore, the values of RQI are almost only located on 50 while $m = 11$, which represents the largest fuzzy existing. Fadili [31] points out that when $m \to 1$, either $\mu_{ik} \to 1$ or $\mu_{ik} \to 0$, the clusters tend to be crisp and while $m \to \infty$, we have $\mu_{ik} \to 1/c$. Hence, the range of $2 \leqslant m \leqslant 5$ seems to be a good compromise for optimising the performance in quality assessment. When $m = 1$, the objective function is the classical within-group sum of squared error (WGSSE), and the μ_{ik}'s can only take the value 0 or 1. Furthermore, when $m = 1$ the partition is hard, and for $m > 1$, the partition is fuzzy. Increasing m causes the partition to become fuzzier [27,28,31].

The curves of RQI for the 101 synthetic data sets, based on five specific quality levels with five varying fuzzy index m, are shown in Fig. 4. Inversion takes place when the subset of synthetic data $\vec{f}_k$ is closed to the assigned standard quality levels, especially when $m = 3$, 5, and 21, where the RQI declines with increasing quality or rises with decreasing quality. This incompatible phenomenon occurs because the methodology changes to crisp theory from fuzzy theory when data points are coincident with any specific quality level. That is, when the observation $\vec{f}_k$ is located coincident with $\vec{e}_i$, then $(\|\vec{f}_k - \vec{e}_i\|^2)^{1/(m-1)} = 0$, and the values of the μ_{ik}s
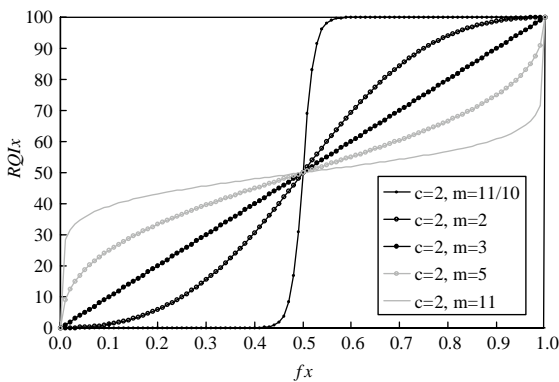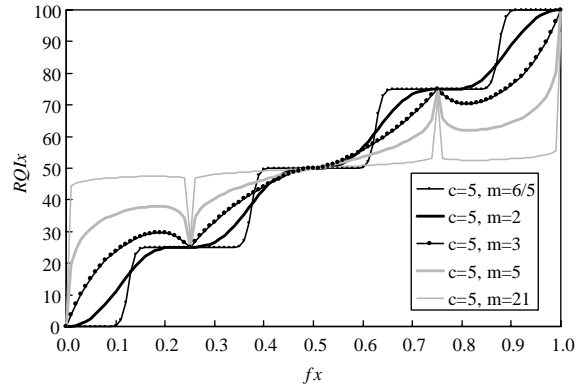

Fig. 4. The curves of RQI for the 101 synthetic data sets based on five specific quality-ordered levels ($c = 5$) with five varying fuzzy index (m).

only go to 0 or 1. Many numbers of specific quality-ordered levels bring about the conflict repeatedly and lead to the contrary. Moreover, for numbers of specific quality-ordered levels, the optimal range of fuzziness index, m, would be limited. The $m = 2$ seems to be acceptable only for optimising performance of quality assessment. Previously, there has been no theoretical basis for an optimal choice for the value of m in the C clustering algorithm; conventionally, $m = 2$ is selected [31,32]. The properties of μ_{ik} for the convergence theory of FCM algorithm applying on quality assessment are explored. The redundant specific quality-ordered levels are unnecessary since they weaken the validity measure of RQI.

3.3. Sensitivity analysis

The sensitivity analysis of the RQI is inspected in this study. From the above discussion, fuzziness index $m = 2$ and two standard quality levels, perfect "bad" and perfect "good", are employed in the river quality assessment for optimising performance. The values of the RQI and the conventional RPI are converse for an observation since the RPI describes river pollution index, but the RQI denotes river quality index. A saturated polluted measurement would produce 0 for RQI and 10 for RPI; on the contrary, an absolute excellent measurement would come out 100 for RQI and 1 for RPI. Thus, for easy comparison between the two indices, the RPIt has been designed. The RPIt ranges from 0 to 9 according to the river quality from bad to good, which is obtained by subtracting 10 from the value of RPI. Hence, both of the RQI and RPIt are quality indices for river bodies.

In order to get the specific transferred vector, the vector of the observation is set with $\vec{x}_k = (x_d, x_b, x_s, x_n) = (4.37, 6.98, 54.81, 1.35)$, where $\vec{f}_k = = (f_d, f_b, f_s, f_n)$

Fig. 3. The curves of RQI for the 101 synthetic data sets based on two specific quality-ordered levels ($c = 2$) with five varying fuzzy index (m).

Table 4
Relative sensitivity analysis of the RQI and the RPIt for the referred samplings (μ_0 = similarity between sample and absolute "bad" quality, μ_1 = similarity between sample and complete "good" quality)

Concentration (mg/L)					Transferred value				Similarity degree		RQI		RPIt	
DO	BOD₅	NH₃-N	SS	Relativity change rate of SS (%)	DO	BOD₅	NH₃-N	SS	μ_0	μ_1	RQI	Relativity change rate (%)	RPIt	Relativity change rate (%)
(x_d)	(x_b)	(x_n)	(x_s)		(f_d)	(f_b)	(f_n)	(f_s)						
4.37	6.98	1.35	19.61	−50	0.5	0.5	0.5	1.0	0.30	0.70	69.8	20	5.25	14
4.37	6.98	1.35	24.01	−40	0.5	0.5	0.5	0.9	0.33	0.67	67.33	20	4.75	14
4.37	6.98	1.35	28.41	−30	0.5	0.5	0.5	0.8	0.36	0.64	63.85	19	4.75	8
4.37	6.98	1.35	32.81	−20	0.5	0.5	0.5	0.7	0.40	0.60	59.68	10	4.75	8
4.37	6.98	1.35	42.71	−10	0.5	0.5	0.5	0.6	0.45	0.55	54.73	4	4.75	8
4.37	6.98	1.35	54.81	0	0.5	0.5	0.5	0.5	0.50	0.50	49.80	0	4	0
4.37	6.98	1.35	66.91	10	0.5	0.5	0.5	0.4	0.55	0.45	44.88	−4	4	0
4.37	6.98	1.35	76.81	20	0.5	0.5	0.5	0.3	0.59	0.41	40.64	−10	4	0
4.37	6.98	1.35	84.51	30	0.5	0.5	0.5	0.2	0.64	0.36	36.34	−19	4	0
4.37	6.98	1.35	92.21	40	0.5	0.5	0.5	0.1	0.67	0.33	32.75	−20	4	0
4.37	6.98	1.35	101.01	50	0.5	0.5	0.5	0.0	0.70	0.30	29.88	−20	3	−11

= (0.5, 0.5, 0.5, 0.5). Each of the four criteria parameters is varied over its possible transition interval, respectively, while the values of the others are kept fixed at the initiation. To explain the relative analysis, Table 4 summarizes the salient features of the changes of suspended solids (SS). Based on the initial vector, the concentrations of SS are changed gradually from 54.81 to 19.61 mg/L, and from 54.81 to 101.01 mg/L while the other criteria are fixed. The similarity degree between the object and the perfect "bad" quality level μ_0 declines with the lower concentration of SS, while the similarity degree between the object and the perfect "good" quality level μ_1 increases. Both the RPIt and RQI decline with the lower concentration of SS. The relative changes of RPIt are from +14% to -11% and RQI from +20% to -20% under the conditions of SS relatively changing from –50% to +50%. The relative changes of RPIt and RQI are almost the same for BOD and NH₃-N. The relative changes of RPIt are from –11% to +14% and those of RQI are from –20% to +20% for DO under the concentration based on 4.3716 mg/L varying from –50% to +50%. The relative sensitivity of the conventional index (RPIt) and the proposed index (RQI) are explicated in Fig. 5. The result shows that three and four staircase steps appear in the conventional index, which is not continuous relating to the change of variable. The relative changes of RQI for the four criteria variables match with their membership functions separately. Furthermore, the overall relative change rate of RQI is greater than that of RPIt.

In summary, the variations of the four criteria parameters are more sensitively represented with RQI than RPIt. The RQI could sharply distinguish the change of quality with the variation of parameters, which is crucial in evaluating quality.

3.4. A case study

The river quality indices of RPIt and RQI are applied for three selected points along the *Keelung River*, the second largest river in Taipei. A map of the area is shown in Fig. 6. The regular monitoring stations of *Nuanjiang Bridge*, *Jiangbei Bridge,* and *Baiyi Bridge* are located upriver, middle-stream, and downstream, respectively. The historical data from 1991 to 2000 are collected from Environmental Protection Administration of Taiwan. Monthly averages are used. There are a total of 120 data records.

A moving average is commonly used to lead the changing tendency of data marked. A 12-month moving average is employed in the case of water quality. The monthly time series and 12-month moving average of RPIt and RQI from 1991 to 2000 are drawn in Fig. 7. The index of RQI shows a remarkable difference among the upriver, middle-stream, and downstream locations. In *Nuanjiang Bridge*, the RQI score is above 80 for the
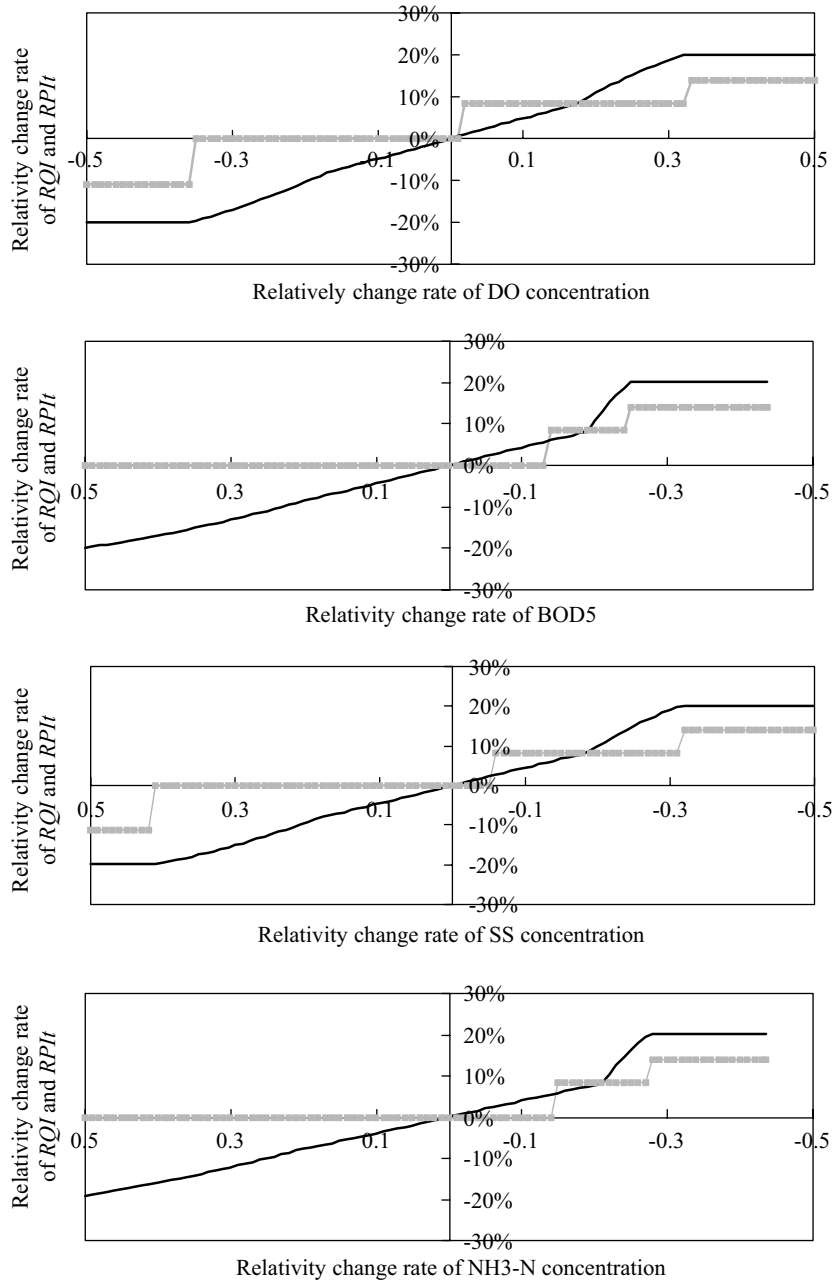
Fig. 5. The relative sensitivity analysis of *RQI* (solid black line) and *RPIt* (dotted gray line).

past 10 years. There is a decrease in quality from *Nuanjiang Bridge* to *Jiangbei Bridge* with a score of 80–40 in the last decade reflecting the effect of the metropolitan area surrounding *Keelung City*. After the river flows into Taipei City, the river is in a seriously polluted state. The score decreases to around 20 in *Baiyi Bridge* for the past 10 years. The discharge of municipal

wastewater is considered to be the primary source of pollution.

The quality of *Nuanjiang Bridge* is excellent since it is located at the source of the *Keelung River*, which has unspoiled and picturesque upper reaches with a number of waterfalls. The quality of the middle-stream of *Keelung River*, *Jiangbei Bridge*, is slightly lower and
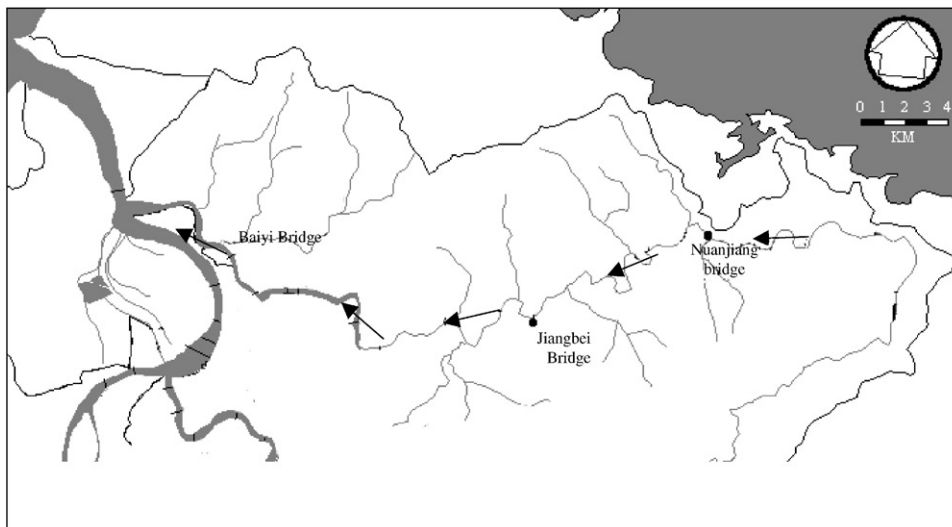
Fig. 6. The spatial configuration of the monitoring station for *Keeling River*.

has fluctuated in the latest decade but seems gradually to have stabilized in the last 3 years. The grossly polluted downstream has seen some gradual improvement from 1998, presumably due to two important remediation events. One of them is the opening of *Bali* sewage-treatment works in 1998. The other relates to a 1990s project that straightened two bends in the river near *Tachi*, which resulted in the conversion of riverbank areas into parkland from the original agriculture areas.

4. Conclusion

In this research, a new and effective gradation model, the two fuzzy set theory, for diagnosing river quality has been developed and illustrated with the case study of the *Keelung River* in Taiwan. A simple numerical scale relating to degree of quality would seem a feasible approach to assess variations in water quality and to convey findings in a comprehensive manner to others [3]. Firstly, environmental monitoring parameters measurements are processed with membership functions relating the various levels of parameter estimates to the appropriate levels of environmental quality. Secondly, the similarity degrees, deriving from the extended convergence theory of the FCM, are weighted, accumulated, and eventually converted into the quality index. Fuzzy theory provides a method that permits an investigator to determine how much a particular set of monitoring measures represent elements of good quality as well as elements of bad quality. *Fonck, Hammah, and Curran* point out that similarity measures between sets are widely used for querying in fuzzy knowledge bases

[29,33]. The model proposed in this research is a new creative idea in environmental evaluation index. It provides a less subjective, more sensitive, and more efficient model for evaluating quality and changes in quality.

From the above discussion, the following conclusions can be drawn.

1. This paper has successfully presented a strategy for the assessment of quality using a similarity degrees method based on the extended convergence of the FCM algorithm.
2. From the study, it is unnecessary to provide redundant specific quality-ordered levels in proceeding consecutive tendency analysis with the extended FCM evaluating model. Moreover, the range of $2 \leqslant m \leqslant 5$ seems to be a good compromise for optimising the performance of consecutive tendency analysis for two standard quality levels, perfect "bad" and perfect "good" being employed.
3. It is obvious that the value of the proposed overall river quality index, RQI, has a linear relationship with the change of the observation compared to the conventional index, RPI. Hence, the index of RQI shows remarkable difference among the upriver, middle-stream, and downstream for the case study of *the Keelung River*.
4. The model proposed in this research is a creative new idea in environmental evaluation. The choice of parameters (fuzziness index m and specific quality-ordered levels c) enhances the freedom of decision-makers. The flexibility of the extended FCM model can engage different evaluating scenarios in quality assessment.

Nuanjiang Bridge
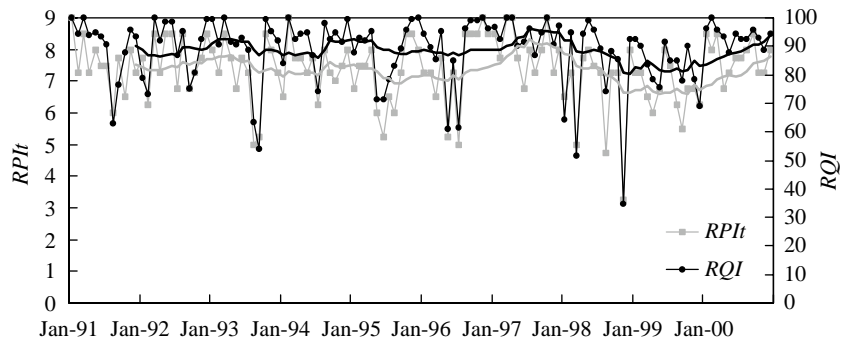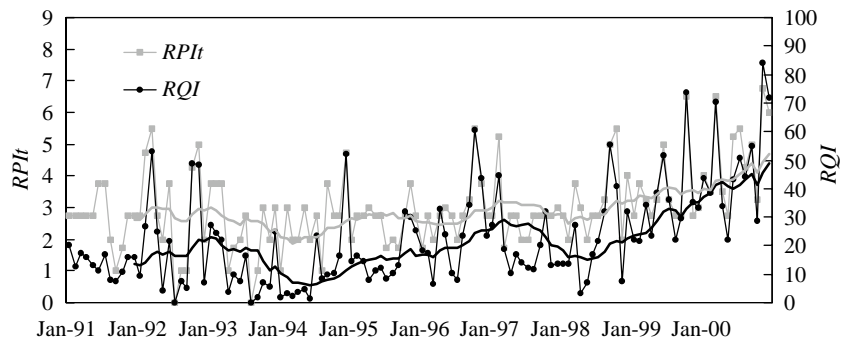

Jiangbei Bridge
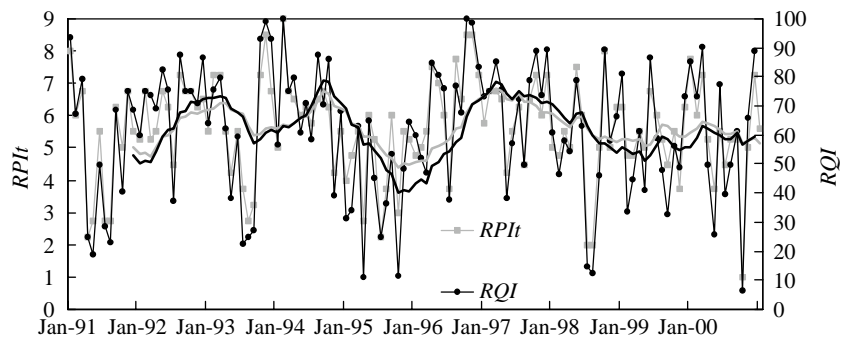
Baiyi Bridge

Fig. 7. The temporal changes of monthly time series (thin line) and 12-months moving average (thick line) for RPI*t* and RQI.

References

[1] Hammond A, Adriaanse A. Environmental indicators: a systematic approach to measuring and reporting on environmental policy performance in the context of sustainable development. World Resources Institute, 1994. pp. 1–50.

[2] Hardi P, Barg S. Measuring sustainable development: review of current practice. Industry Canada Occasional Paper No. 17, 1997. pp. 1–27.

[3] Ross SL. An index system for classifying river water quality. Water Pollut Control 1977;76(1):113–22.

[4] Landwehr JM. A statistical view of a class of water quality indices. Water resour res 1979;15(2):460–8.

[5] Ball RO, Church RL. Water quality indexing and scoring. J Environ Eng Div ASCE 1980;106(4):757–71.

[6] House MA, Ellis JB. The development of water quality indices for operational management. Water Sci Technol 1987;19(9):145–54.

[7] House MA, Newsome DH. Water quality indices for the management of surface water quality. Water Sci Technol 1989;21:1137–48.

[8] Smith DG. A better water quality indexing system for rivers and streams. Water Res 1990;24(10):1237–44.

[9] Vega M, Pardo R, Barrado E, Deban L. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Res 1998;32(12):3581–92.

[10] Nives SG. Water quality evaluation by index in Dalmatia. Water Res 1999;33(16):3423–40.

[11] Swamee PK, Tyagi A. Describing water quality with aggregate index. J Environmental Engineering 2000;126(5):451–5.

[12] Helena B, Pardo R, Vega M, Barrado E, Fernandez JM, Fernandez L. Temporal evaluation of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. Water Res 2000;34(3): 807–16.

[13] Nagel JW. A water quality index for contact recreation. Water Sci Technol 2001;43(05):285–92.

[14] Silvert W. Fuzzy indices of environmental conditions. Ecol modeling 2000;130:111–9.

[15] Zadeh LA. Fuzzy sets. Inf and Control 1965;l(8):338–53.

[16] Genther H, Glesner M. Advanced data preprocessing using fuzzy clustering techniques. Fuzzy Sets and Systems 1995;85:155–64.

[17] Shivathaya SS, Fang XD. Fuzzy-logic-based ranking of steel grades generated by material design KBS. Eng Appl Artif Intell 1999;12:503–12.

[18] Kung HT, Ying LG, Liu YC. A complementary tool to water quality index: fuzzy clustering analysis. Water Res Bull 1992;28(3):525–33.

[19] Hayo MG. Zimmer C. An indicator of pesticide environmental impact based on a fuzzy expert system. Chemosphere 1998;36(10):2225–49.

[20] Tao Y. Fuzzy comprehensive assessment: fuzzy clustering analysis and its application for urban traffic environment quality evaluation. Transp Res 1998;3(1):51–7.

[21] Wang WJ. New similarity measures on fuzzy sets and on elements. Fuzzy sets and Systems 1997;85:305–9.

[22] Bezdek JC. Fuzzy partition and relations: an axiomatic basic for clustering. Fuzzy Sets and Systems 1978;1:111–27.

[23] Dee N. An environmental evaluation system for water resource planning. Water resour res 1973;9(3):523–35.

[24] Windham MP. Geometrical fuzzy clustering algorithms. Fuzzy Sets and Systems 1983;10:271–9.

[25] Bezdek JC. A convergence theorem for the fuzzy ISODATA clustering algorithm. IEEE Trans. Pattern Anal. Machine Intell. 1980;PAMI-2(1):1–8.

[26] Bezdek JC, Hathaway RH, Sabin MJ, Tucker WT. Convergence theory for fuzzy c-Means: counterexamples and repairs. IEEE Trans. Sys Man Cybern 1987;17: 873–7.

[27] Flores-Sintas A, Cadenas JM, Martin F. Membership functions in the fuzzy c-means algorithm. Fuzzy sets and system 1999;101:49–58.

[28] Rezawee MR, Lelieveldt BPF, Reiber JHC. A new cluster validity index for the fuzzy c-means. Pattern Recognition Letters 1998;19:237–46.

[29] Hammah RE, Curran JH. One distance measures for the fuzzy k-means algorithm for joint data. Rock mech. Rock Eng 1999;32(1):1–27.

[30] Silvert W. Fuzzy indices of environmental conditions. Ecological modeling 2000;130:111–9.

[31] Fadili MJ. On the number of clusters and the fuzziness index for unsupervised FCA application to BOLD fMRI time series. Med Image Anal 2001;5:55–67.

[32] Klawonn F, Annette K. Fuzzy clustering based on modified distance measures. IDA'99, Lecture Notes in Computer Science, Vol. 1642. Brelin: Springer, 1999. p. 291–301.

[33] Fonck P, Fodor J, Roubens M. An application of aggregation procedures to the definition of measures of similarity between fuzzy sets. Fuzzy sets and Systems 1998;97:67–74.